DOCUMENT RESUME

ED 273 205 HE 019 620

AUTHOR Scheetz, James P.
TITLE Some Perspectives on the Use of Student Ratings to Evaluate Teaching Effectiveness.
PUB DATE Jul 86
NOTE 6p.
PUB TYPE Viewpoints (120) -- Journal Articles (080)
JOURNAL CIT Professions Education Research Notes; v8 n1 p4-6,9 Jul 1986

EDRS PRICE MF01/PC01 Plus Postage.
DESCRIPTORS *College Faculty; College Instruction; College Students; Dental Schools; *Faculty Evaluation; Higher Education; Student Attitudes; *Student Evaluation of Teacher Performance; *Teacher Effectiveness
IDENTIFIERS University of Louisville KY

ABSTRACT

Issues concerning the development of a teaching evaluation system are considered, along with the use of student ratings of teachers. In addition to briefly reviewing the beginnings of teacher evaluation, the following prerequisites for a successful faculty evaluation system are identified: the cooperation and support of many groups, strong administrative support, faculty involvement, and expert assistance. Three dimensions of teaching effectiveness are distinguished: (1) the structure, organization, and clarity of courses and classroom presentations; (2) student-teacher interaction; and (3) communication ability of the teacher. Research evidence related to student ratings is summarized, and the evaluation system used by the University of Louisville School of Dentistry is described. This system involves course evaluations and student ratings of faculty. A range of activities are assessed, including the use of objectives, testing and grading procedures, instructor assistance in problem-solving, and instructor's attitude toward dental students and the profession. It is emphasized that student ratings are one of many measures that can be used to assess teaching effectiveness. (SW)

****
* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *
****

ED273205

SOME PERSPECTIVES ON THE USE OF STUDENT RATINGS TO EVALUATE TEACHING EFFECTIVENESS

James P. Scheetz
University of Louisville

"PERMISSION TO REPRODUCE THIS MATERIAL HAS BEEN GRANTED BY

James P Scheetz

TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)."

U.S. DEPARTMENT OF EDUCATION
Office of Educational Research and Improvement
EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)

This document has been reproduced as received from the person or organization originating it.

☐ Minor changes have been made to improve reproduction quality.

• Points of view or opinions stated in this document do not necessarily represent official OERI position or policy.

HE019620

Some Perspectives On the Use of Student Ratings To Evaluate Teaching Effectiveness

James P. Scheetz
University of Louisville

Faculty evaluation often occurs in a setting which can lead to feelings of antagonism and distrust. With the current focus on accountability and the measurement of outcomes, it is unlikely that efforts to measure faculty effectiveness will decrease. It is much more likely that assessment procedures will become more prevalent with feelings by the people being evaluated that these procedures are somewhat unfair and the results are used in a punitive fashion. It is not surprising that faculty members may be resistant to attempts to assess their effectiveness and productivity.

In the area of scholarship and research, effectivenss can be evaluated by counting the number of journal articles published and the presentations at meetings made by a faculty member. In addition, the amount of grant money secured by an individual is relatively easy to determine. Although these mechanisms may be viewed as relatively crude measures of faculty productivity in terms of scholarship and research, many faculty members and administrators use these measures with little disagreement that these activities are indicators of productivity and effectiveness.

When measuring faculty service, there are several traditional measures which are employed. Among these are participation in professional organizations, service to the employing institution (e.g. serving on committees), and community service. Again, these may be viewed as relatively crude measures, but they are not difficult to use.

Another area in which performance of faculty is assessed is teaching. In contrast to scholarship and research and service, the assessment of teaching effectiveness presents complex problems. First, it is more difficult to use a counting procedure as can be used with scholarship and service when evaluating teaching. Second, it may be difficult to formulate a definition of which activities constitute effective teaching. Third, even if a definition of effective teaching is agreed upon, measurement problems may arise in developing mechanisms for evaluating teaching effectiveness.

The discussion which follows will focus on some of the issues which arise in developing a teaching effectiveness evaluation system. Further, some experiences which occurred during the implementation of such a system will be discussed with the purpose of exploring possible problems in using a teaching evaluation system.

Prerequisites for a Successful Faculty Evaluation System

Faculty evaluation systems require cooperation and support of many groups. McNeil (1981) has indicated that four major conditions must be present for an evaluation system to be successful. First, a recognized need must exist for the establishment of a system or for revision of an existing system. This need must be pervasive because implementation of an evaluation system requires the cooperation of many groups and individuals.

Second, strong administrative support is required. If the administration believes an evaluation system is beneficial it is more likely that the resources needed to make a system function smoothly and efficiently will be provided. Many times the stimulus for development of an evaluation system will come from the administration as the importance of excellence in teaching is stressed.

Third, faculty involvement is essential in implementing an evaluation system. If the faculty are opposed to being evaluated, they may subvert the efforts to evaluate their performance. For this reason, the administrators involved in this activity should be sensitive to faculty opinion. If it becomes evident that the faculty are opposed to the operation of an evaluation system, extreme caution should be exercised. If the faculty strongly resist the attempt to evaluate teaching, it is unlikely that the goal of implementing a system can be accomplished. Securing faculty support and cooperation, even if a relatively long period of time is required, should lead to faculty acceptance of the system.

Fourth, expert assistance is usually required in the development and revision of faculty evaluation systems. Many times the individuals who make the decision that teaching ability should be assessed are not knowledgeable about developing assessment mechanisms. A poorly developed system may be a target of faculty complaints resulting in the faculty having little faith in the system. For this reason, individuals with skills in developing evaluation procedures should have a major role in the development of the evaluation program. The development process may be time consuming and require much effort by the people involved, but the investment of time and effort usually results in a system with relatively few flaws which must be corrected at a later time.

The cooperation of another group not mentioned by McNeil is also essential; the students enrolled in courses. In much of the literature, student participation is treated as a given, but the same careful attention which is focused on faculty development shouls be focused on student involvement. Without student cooperation, a faculty evaluation system is not workable.

In summary, the cooperation and efforts of many individuals are required to design and implement a successful faculty evaluation system. This should be recognized by anyone involved in this type of activity. In the next section a brief review of the beginnings of teacher evaluation will be presented.

How We Got Where We Are with Regard to Student Ratings

Prior to the 20th century, it was generally agreed that students were responsible for their own learning and faculty or teachers had little or no responsibility. Teachers were viewed primarily as managers of learning activities who provided the opportunity for learning. However, if the students did not learn the assigned material the teacher was not held responsible. Travers (1981) stated that teachers were not held accountable except for management problems which might impede student learning. Thus, teachers were expected to listen to students' recitations and provide new assignments.

When it became common practice to assign students to grades according to age groups, the role of the teacher as primarily a manager of learning activities began to change. When all students in a classroom were assigned the same material, the teacher could then address the problems which the students encountered in completing their assignments. This led to a more active instructional role for the teacher, such that the performance of the teacher as well as that of the students was evaluated. Teacher evaluations were accomplished primarily by having a person in an administrative position visit the classroom. This person could assess both the instructional ability of the teacher and the teacher's method of maintaining discipline.

At the present time some people involved in management of educational programs assume that the students have little or no responsibility for learning. If students fail to learn, the teacher is held accountable and is blamed for providing inadequate conditions for learning to occur. This view absolves the students of any responsibility if they do not learn the assigned material. This conception of the educational process represents a drastic change from the view held prior to the 20th century that students were solely responsible for their

own learning. If the teacher is held totally accountable for student learning, it is to be expected that teacher performance will be evaluated in minute detail. Elaborate schemes have been developed to accomplish these evaluations. Merit pay for teachers has been proposed as a method for improving the quality of instruction. If a merit pay system is implemented, teacher performance must be measured. This is one example of a system which may lead to further emphasis on evaluating teacher performance.

If the current view - that teachers are responsible for student achievement - prevails, then emphasis on teacher evaluation will be likely to continue. Perhaps a more realistic view of the educational process is that both teachers and students are responsible for the level of student achievement. If this view is accepted, less emphasis may be placed on teacher evaluation and a more balanced perception of the teaching-learning process may emerge. The educational community may benefit if less emphasis is placed on evaluation so that teachers can try innovative techniques without fear of the consequences associated with an unfavorable evaluation.

If teaching efforts are to be assessed, what should be evaluated? In the next section, criteria of effective teaching will be explored.

Criteria of Effective Teaching

If teacher ability is to be evaluated, it is imperative that the qualities being assessed are clearly specified. Centra (1979) has summarized the findings of several studies. He presents several skills as being important for effective teaching.

If a classroom presentation is to be effective, it must hold the interest of the students. The speaking ability of the teacher is related to student interest in the material being presented. Teaching is judged to be more effective if the teacher speaks audibly and clearly so that the students do not have to strain to hear what is said.

Classroom presentations which are well organized are more effective than those in which the material is presented in a haphazard fashion. If the teacher summarizes or emphasizes major points it is easier for the students to follow the presentation. Lack of organization may lead students to perceive instruction as somewhat ineffective.

The rapport which the teacher has with students influences student assessment of instruction. Good rapport can lead to student enthusiasm resulting in more student participation and receptivity to learning. The teacher who ignores this aspect of the teaching-learning process is ignoring an important aspect of instruction.

Facilitating group interaction is one means of encouraging students to become more active in the learning process. Those teachers who recognize this and encourage interaction are usually judged to be more effective than those who do not. Teachers who expect passive rather than active student participation are depriving students of an avenue of learning which can be beneficial.

Those teachers who are responsive, accessible and interested in students project a willingness to become involved with students to enhance their learning. This willingness encourages students to ask questions and seek help from the teacher if needed. This activity leads students to judge those individuals to be effective teachers.

If the teacher demonstrates enthusiasm about the subject matter and teaching, this enthusiasm may carry over to the students so that they become more attuned to the instruction presented by the teacher. Anything that increases student interest will usually lead to a more effective learning environment. Teachers who are enthusiastic are more able to motivate students to become active learners.

Along with demonstrating enthusiasm about the subject matter area, the teacher who is knowledgeable about the subject will be seen as more effective than a teacher who lacks or does not demonstrate this knoweldge. This suggests that as far as possible teachers be assigned to teach those courses in which they are content experts. Lack of knowledge about the subject matter influences students' judgment of teacher competence.

Teachers who are flexible in their approach to teaching are more effective than those who are dogmatic. Teacher flexibility indicates a willingness to deviate from planned activity thereby indicating a sensitivity to student input. Teachers who can bend to student suggestions and still retain control can stimulate student learning through this exchange of ideas.

In summary, three overall dimensions of teaching effectiveness have been identified. First, the structure, organization and clarity of courses and classroom presentations are important. Second, teacher effectiveness is influenced by student-teacher interaction or rapport. Third, teaching skills encompassing communication ability influence student perceptions of teacher effectiveness.

If student ratings are one method of assessing teaching ability, what are the strengths and limitations of these ratings? In the next section some current views about student ratings will be summarized.

Research Evidence Related to Student Ratings

Centra (1979) concluded that small classes (15 or less) are rated higher than larger classes. Small class size is usually more conducive to interaction between the teacher and the students. It is also easier to adapt the level of instruction to the needs of the students. The importance of interaction is related to ratings in that discussion courses tend to receive higher ratings than lecture courses. Thus, interaction between the teacher and students, and among students, may lead to more favorable student ratings.

Students generally rate elective courses and courses in their major area of study higher than they rate other courses. Centra (1979) speculates that students are more motivated and interested in these courses than in other courses. Students may view elective and major field of study courses as more relevant to their future employment which could lead to higher ratings. However, the evidence on the relationship between course type and ratings is somewhat tentative.

It has been argued that only after individuals have completed a program of study can they evaluate courses in which they were enrolled. This argument leads to exaiming the relationship between student and alumni ratings. In general, the correlation between student and alumni ratings is relatively high. Both Centra (1979) and Aleamoni (1981) have concluded that ratings of teachers and courses change very little over time when follow-up studies of graduates' ratings are obtained.

One of the most controversial topics related to student ratings is the relationship between grades and ratings assigned by students. Centra (1979) has stated that the correlation between assigned grades and ratings is approximately 0.20. He reported that the correlation between expected grades and ratings is slightly higher. Stumpf and Freedman (1979), in an exhaustive survey of the literature, reported that correlations between grades and ratings ranged from -0.75 to 0.75. The relationship between grades and ratings, as revealed by correlational measures, appears to be mixed. It is possible to find evidence of both positive and negative relationships between grades and ratings.

In an experimental study, Chacko (1983) investigated the effect of grades on student ratings. The treatment consisted of the instructor's evaluation of student performance on an examination. The findings of this study indicate that student ratings are influenced by instructor evaluation of student performance. This suggests that ratings should be obtained before students are informed of the instructor's evaluation and before grades are known by students. Chacko states that ratings may be interpreted as emotional reactions implying less objective evaluations by students than would be required for credible ratings. He concludes:

The findings provide rather strong, indictments against the use of student ratings as a means of obtaining unbiased information.

In this section the evidence related to the use of student ratings of instruction has been summarized. The next section deals with the evaluation system used in the University of Louisville School of Dentistry that is composed exclusively of student evaluations.

The Dental School Evaluation System

The School of Dentistry evaluation system focuses on both course and faculty evaluation. Since many courses involve participation on the part of several faculty members, it is necessary to separate course and faculty evaluation. In addition, the format of the course may include several instructional settings including lectures, seminars, laboratory instruction and clinical instruction. Aleamoni (1981) has described this type of system as one in which multiple standard forms are available. The form used may vary according to the mode of instruction, but the items on each form are constant.

At the end of each semester the students in each of the four classes meet for approximately 30 minutes to complete the evalaution forms. Every course is evaluated and selected faculty are rated by the students. Given the large number of faculty involved in many of the courses, it is unrealistic to ask the students to rate every faculty member who has instructed them in a particular semester.

Each rating form contains 12 to 14 items which are specifically related to the mode of instruction for the instructor evaluation forms. There are two course evaluation forms; one for didactic and one for clinical courses. The items assess a range of activities including the use of objectives, testing and grading procedures, instructor assistance in problem solving, instructor enthusiasm for teaching, and the instructor's attitude toward dental students and the dental profession.

The data obtained from the rating forms is computer scored. The printouts for the course evaluations are distributed to the course directors and departmental chairpersons. The faculty evaluation printouts are given to the individual faculty members, the course director and the chairpersons. Thus, the only people who receive the faculty evaluation forms are the course director, the faculty member and the chairperson. No printouts are distributed to dental school administrators. At this point in time, the results are for feedback and the chairperson uses the results in making personnel decisions.

The responses are analyzed by computing the mean, standard deviation and the median with only the median being reported. The reason for reporting the median is to provide a typical measure of teaching effectiveness. Thus, discussions which focus on the difference between a rating of 3.20 versus 3.28, as might be obtained if arithmetic means are reported, is avoided. In addition, the response frequency is reported.

Another measure of the quality of the evaluation system is the internal consistency measure of reliability, as measured by Coefficient Alpha. Across all of the rating forms, this measure has consistently been above 0.90. This indicates that a unitary construct is being measured. The content validity was implicitly assessed in the developmental stage when the students and faculty agreed to the inclusion of the items appropriate for evaluation of courses and faculty.

The School of Dentistry evaluation system has been briefly described. In the next section the reactions of the author, who coordinated the development of the system, will be explored.

Reactions to Using the Evaluation System

My reactions and thoughts about the system have evolved over several years. One of the first things that became evident is that a bias and halo effect may exist among the students. That is, popular instructors may receive favorable ratings based on their reputation. Conversely, unpopular instructors may be rated harshly by the students, again based on reputation. Since the students in the four classes have frequent contact with each other, there is ample opportunity for the students to discuss instructors and, in my opinion, some students prejudge instructors before a course begins. Just as students may have prior opinions about an instructor, they may form prior opinions about courses. Thus, some course ratings may be colored by the students' preformed ideas about a particular course.

A dilemma faced by some course directors is what to do about course content judged by the faculty to be essential, but criticized by the students. If a course director is overly sensitive to student ratings, he or she may delete content which the faculty believe should be included in the course. There is no ready solution to this dilemma. The course director must exercise caution and judgment in making course changes based on feedback from student ratings.

Students have expressed some impatience with seeing little or no change resulting from their efforts. It is important for them to understand that changes do occur, but these changes occur slowly and an unpopular faculty member, for example, may not be terminated solely because of low student ratings. Students must be made to realize that they have benefited from the changes as a result of the ratings provided by students who have preceded them; and that succeeding classes will benefit if they provide accurate and thoughtful ratings of courses and instructors.

Given the format in which the ratings are obtained with the students rating courses and instructors at a single sitting, fatigue may be a confounding factor. It would be preferable if the ratings for each course and the faculty members who teach the course were obtained as part of the course activities. This would require approximately five to ten minutes in each course and the students might be more inclined to reflect upon and accurately assess the instructional activities in each course. However, it is my opinion that some course directors would refuse to allocate time for collecting this information. This is the primary reason for asking the students to rate courses and faculty in a single sitting.

With regard to the development of the evaluation system, the students were much more enthusiastic than the faculty. This is not surprising in that it provides an opportunity for the students to evaluate those who regularly evaluate student performance. The students were willing to invest their time and effort in developing the system. The faculty reacted much more passively and expressed some skepticisim that such a system was needed and could be workable. However, the faculty did not actively oppose the implementation of the evaluation system.

The reactions of the system coordinator are somewhat personal and without documentation. In the last section, the question of the utility of students ratings is addressed.

Conclusions

Perhaps the ultimate question is: Should student ratings be used to evaluate courses and instructors? In the literature there are arguments both for and against the use of student ratings. In my opinion, the answer to the question is a qualified yes. Student ratings are one of the many potential sources of information about instructional effectiveness. It would be a mistake to judge a course or instructor solely on the basis of student evaluations. A comprehensive evaluation system should include other measures in addition to student ratings.

The current system includes only student ratings, which is a shortcoming. The results are used both for feedback and for making personnel decisions. If the ratings from any system are used for personnel decisions, some safeguards must be included so that a faculty member is not dealt with in a capricious manner. After observing the current system for several years, in particular student reaction and comments while filling out the forms, the results should be approached with extreme caution. Again, it should be emphasized that student ratings are one of many measures which can be used to assess teaching effectiveness.

In conclusion, if we hold the teacher partially responsible for student learning, then it is probably inevitable that teach-

(continued, page 9)

Scheetz (continued)

ers will be evaluated. However, this evaluation should be carried out in a fair and humane manner.

References

Aleamoni, L.M. Student ratings of instruction. In *Handbook of Teacher Evaluation*. J. Millman (ed). Beverly Hills, CA: Sage Publications, 1981.

Centra, J.A. *Determining Faculty Effectiveness*. San Francisco: Jossey-Bass Publishers, 1979.

Chacko, T.I. Student ratings of instruction: A function of grading standards. *Educational Research Quarterly*, 1983, *8*:19-25.

McNeil, J.D. Politics of teacher evaluation. In *Handbook of Teacher Evaluation*. J. Millman (ed). Beverly Hills, CA: Sage Publications, 1981.

Stumpf, S.A. and Freedman, R.D. Expected grade covariation with student ratings of instruction: Individual versus class effects. *Journal of Educational Psychology*, 1979, *71*:293-302.

Travers, R.M.W. Criteria of good teaching. In *Handbook of Teacher Evaluation*. J. Millman (ed). Beverly Hills, CA: Sage Publications, 1981.

Our Esteemed Authors....

Carolyn L. Cason, Ph.D.
University of Arkansas
College of Medicine
4301 West Markham Street
Little Rock, AR 72205
501/661-5163

Gerald J. Cason, Ph.D.
University of Arkansas
College of Medicine
4301 West Markham Street
Little Rock, AR 72205
501/661-5720

Frank T. Stritter, Ph.D.
University of North Carolina,
Chapel Hill
ORDEHP-322 McNider Hall-202H
Chapel Hill, NC 27514
919/966-3641

James P. Scheetz, Ph.D.
University of Louisville
School of Dentistry
Health Sciences Center
Louisville, KY 40292
502/588-5184

PERN Wants YOU!

This is the spot, the means and the method to communicate with your colleagues. Right here, in the pages of *PERN*, you can address those issues you've discussed only with your word processor, find the collaborators you need for a multi-institutional study, discover the wheels that others have already invented for the problem you need to solve. If your aim is to communicate, *PERN* is the place. So send your letters, announcements, or manuscripts. (With all due respect to Abe, no scribbled backs of envelopes, please.)

Letters (continued)

We appreciate Dr. Cason's thoughtful comments on our article. We agree with many of the points he makes, and we regret any confusion our introductory paragraphs may have caused. Perhaps had we used the terms "type X item" in these paragraphs, our intentions would have been clearer. However, our purpose was not to compare type K and type X items, but rather to acknowledge that there is more than one multiple-true false (MTF) item format, and that some difficulties may be encountered in constructing effective MTF items incorporating a response code.

Our article was based on work with faculty members who were writing MTF items using response codes. Dr. Cason has listed several possible reasons for this practice. In our case, the most probable reason was practice for national examinations. In addition, these faculty members were (and still are) essentially subject to the technological limitations he describes. Thus, the purpose of our article was to offer some suggestions which will help faculty members in such circumstances construct more effective test items by avoiding the difficulties identified in our article.

Theodore W. Whitley
East Carolina University

Announcements

New President for AAMC

Dr. Robert Petersdorf, Vice Chancellor for Health Sciences and Dean of the School of Medicine at the University of California, San Diego, on March 26 was appointed President and Chief Executive Officer of the Association of American Medical Colleges. His appointment will be effective on September 2, when he succeeds **Dr. John A.D. Cooper,** who is retiring after 17 years as head of the Washington-based association.

Dr. Petersdorf earned his B.A., cum laude, from Brown University and his M.D., cum laude, from Yale. Before assuming his present posts at San Diego, he was Chairman of the Department of Medicine at the University of Washington for 15 years and President of Brigham and Women's Hospital from 1979 to 1981.

FREE! Now *that's* cost containment!

The W.K. Kellogg Foundation and the National Fund for Medical Education sponsored a national invitational conference in 1984 to deal with teaching and learning as they relate to cost effective health care. Attended by many physicians, noted academicians, health care officials, and leaders of several demonstration projects, the conference has resulted in a new book, **Medical Education: Making the Grade in Cost Containment**. You may reserve a free copy of this book, courtesy of the W.K. Kellogg Foundation, by writing to...W.K. Kellogg Foundation, Communications Office, 400 North Avenue, Battle Creek, Michigan 49017-3398.